Against the Grain

Volume 9 | Issue 3 Article 33

June 1997

Chaos-Standard Identifiers for Electronic Information

Albert Simmonds
R.R. Bowker

Sandra K. Paul
SKP Associates

Chaos

by **Albert Simmonds**

Column Editors: **Sandra K. Paul** (President, SKP Associates) and **Albert Simmonds** (Director, ISBN —US Agency, R.R. Bowker)

Introduction

A large portion of the February 1997 meeting of the **Serials Industry Systems Advisory Committee** (SISAC) was given over to a discussion of how to identify information. In part this discussion centered on the newly revised **Serial Item and Contribution Identifier (SICI).** [ANSI/NISO Z39.56-1996 (version 2), available from NISO Press.] A committee of the **National Information Standards Organization** (NISO) in Washington D.C. had recently completed taking this standard through a lengthy revision process, and the SISAC meeting was seen as an important forum for disseminating information about these revisions. The discussion was also important in that it took place within the context of a larger discussion: the development of standard identifiers for electronic information.

Publishers are now at a crucial stage in their strategic planning in the area of electronic publishing. For publishers of learned journals in the scientific, technical and medical (STM) world, the age of electronic publishing has already begun. A number of these journals have already been put into digital formats, and for both the publishers and their customers the future is clearly turning into the digital age. It is also clear that subsequent development of standard indentifiers for electronic publications must be informed by the needs of the publishers which will have the responsibility for assigning them. This development must also take into account the processes of a wide range of other entities in the information world whose needs will shape how the world of information will operate in the twenty-first century.

Paper Identifiers

In thinking about standard identifiers for electronic information, it is natural that we look first at the identifiers such as ISBN that have served us so well in the paper world. We have come to understand, for example, that ISBN must be unique and persistent. We also require that our identifiers be international in scope in order to match the international structure of the information industry. And there must be a large enough pool of identifiers to number many objects, as well as many kinds of objects. (In the United states alone, 135,000 new ISBNs are recorded annually and 7500 new prefixes are requested each year.) Because these characteristics are inherent in ISBN, and because of the successful management of this identifier within an international system, the ISBN has proven a success in numbering monographs published in paper formats, just as the ISSN successfully numbers journal publications. For example, the EAN bar code, used to encode the ISBN for printing on book jackets and covers, can be read by scanners around the world, and users of this bar code can be certain of a high degree of accuracy in identifying the publication scanned.

Minimum Requirements for Identifiers

There are, then, certain minimum requirements for identifiers of publications: uniqueness, persistence, international applicability. And there must be a large enough pool of identifiers to keep pace with the number of publications to be identified. This last point looms as an even greater challenge for identifiers in the electronic age. Current thinking is that the output of electronic information to be identified will increase exponentially over current output in the largely paper world of twentieth century publishing. In addition, the users of electronic documents will probably not be content with an identifier that notes the work itself but will also want to identify the component parts of a work.

Dr. Norman Paskin, Director, Information Technology Development at **Elsevier Science Ltd.**, points out the following:

"At the highest level, an obvious item that can be identified in a library environment is a cluster of documents represented by some physical manifestation: a particular journal, a book containing several contributions, or even a database. That cluster can be given an identification label of some sort (e.g., an ISSN for a particular journal). Within that cluster, there will be the second level, individual documents; these may be single chapters in a book, individual papers written by a scientific author in a journal, or some other form: one can envisage a map or a compilation of data being treated as a document ... it may well become possible and useful to separately identify as a third and lowest level components of a document, since in a hypertext environment it is easy to conceive of links to individual tables within a document, or the same figure used in several different documents. In this article I discuss the level of the single document, but the context of other levels above and below is relevant, and one should bear in mind the wider world of multimedia objects, likely to become of more significance." [See "Information Identifiers in *STM Newsletter* 101.]

Thus, we are looking at a future in which we must have a rigorous identification scheme for the component parts of electronic books and journals — all those parts that need to be identified for either commerce or rights management. This point is implicit in the revisions to the SICI, which now, according to NISO's introduction to the standard "defines a variable length code that will provide unique identification of serial items (e.g., issues) and the contributions (e.g., articles) contained in a serial title."

Differences Between Print and Electronic

We also need to account for the fact that the differences between books and journals in the electronic environment are much less pronounced than in the world of paper publications. If we need a SICI, then we also need a BICI. And, in fact, the **Book Item and Content Identifier** is currently being developed as a joint project of Book Industry Communication in the U.K. and NISO. The purpose of this embryonic standard will be to provide unique, persistent identification of the component parts of ISBN publications. As currently envisioned, the BICI will very closely follow the formulations of the revised SICI.

Another important point in the evolution of electronic publishing (alluded to in the last sentence of the above quotation from Dr. Paskin) will be the growth of multimedia works. Text will mingle freely with music and images in twenty-first century publishing. The importance of this issue has been recognized by the formation of a working group by the International Standards Organization, which is charged with the Coding of Moving Pictures and Audio. The members of this group are currently working with representatives of the information industry to provide standardized technical solutions for requirements in the following areas: • administration of identifiers; • protection of content (protection of access to works); • report on the use of works (logging of manipulation by the user); • trans-

continued on page 79

Chaos
from page 78

actions between users and rights administrators.

Providing the liaison between this ISO group and the information industry is **M. Dominique Yon** of the **International Confederation of Societies of Authors and Composers (CISAC)** in Paris.

DOI

Another important development is represented by the **Digital Object Identifier** initiative of the **Association of American Publishers** which was demonstrated at the AAP's Professional and Scholarly Publishers' meeting in February. As noted by the AAP: "The DOI System connects customers, via the Internet or any future networks, with the current copyright holder of a digital object. Its purpose is to facilitate digital commerce and scholarly research by ensuring the uninterrupted availability of publishers' digital objects over time."

The technology underlying the DOI system is called the **Handle System**. It was developed by the **Corporation for National Research Initiatives (CNRI)**, an internationally recognized leader in information technology research and development located in Reston, Virginia. Since its inception in 1986 as a non-profit organization, CNRI has been actively engaged in the establishment of open, non-proprietary technological approaches for networked systems. The president of CNRI is **Dr. Robert E. Kahn**, who spent thirteen years at the **Advanced Research Projects Agency (ARPA)** before founding CNRI. The DOI system development team at CNRI is headed by **Constance McLindon**.

A prototype of the system is currently under development, and the status of this prototype will be demonstrated at the **Frankfurt Book Fair** this fall. Videotapes of the February demonstration are available from the AAP's Washington office (1718 Connecticut Ave., Suite 700, Washington, D.C., 20009-1148: $15 US in NTSC format and $20 US in SECAM and PAL formats.) In addition, information can be found at <http://www.doi.org>.

Coming Soon

At the next meeting of the **International Standards Organization (ISO)** in London a delegation from NISO will lead a discussion on standards for electronic identifiers. This discussion promises to further our understanding of the framework within which electronic identifier standards will develop. Your editors from *ATG* plan to publish a report on this meeting in the September issue.

Innovations
from page 76

to agree on a copy protection solution. The issue revolved around the movie industry's decision not to provide content without such protection. The hardware and software producers, on the other hand, were not as concerned with copy protection as with developing new markets and expanding existing ones.

The Motion Picture Association of America and the **Consumer Electronics Manufacturers Association** proposed copyright legislation, called the **Digital Video Recording Act** in April, 1996. That act intended to enforce compliance with a specific method of copy protection. It would have also made any importation, manufacture, or distribution of devices that do not comply with the stated method of copy protection punishable by law. That proposal met with much opposition and had to be re-worked.

An ad hoc group of 90 representatives from 60 companies and trade groups, known as the **Copy Protection Technical Working Group**, finally agreed on a proposal for a worldwide system for protecting material at a meeting in mid-September, 1996. Instead of endorsing a single manufacturer's method, the participants made the protection scheme part of the DVD specification. They agreed to have different encoding formats for each of six different geographical regions, similar to the NTSC and PAL video formats. This would make it difficult to copy or play discs created in a region other than that of origin. For example discs made in the U.S. will not play on drives produced in Europe or Asia.

The copy protection agreement requires every DVD-ROM drive and every MPEG-2 card to have a method that will protect encrypted DVD movie discs from being copied. Its only purpose is to prevent people from watching or copying movies on a PC and to force them to pay for extra hardware which they don't need in most cases. It has no effect on other types of data, other than encrypted and scrambled audio and video. Software publishers and owners or providers of textual and still graphic information will need to find another way to protect their products.

System Availability

Shortly after reaching agreement on the copy protection issue, **Toshiba** and **Matsushita** began selling DVD players in Japan on November 1, 1996. **Akai Electric Co., Ltd.** quickly followed suit. **Pioneer** began selling DVD players in the U.S. on March 1, 1997. Toshiba planned to delay introducing DVD players in the U.S. until the end of the first quarter of 1997, while California-based **Diamond Multimedia** announced plans to market Toshiba's drive to end-users in a DVD multimedia upgrade kit in early 1997. The kit includes Diamond's PCI-based adapters which include decoders for MPEG-2 video and Dolby Digital Surround Sound AC-3 audio and copy protection decryption hardware. **Creative Labs, Inc.** also developed a new line of DVD products for the PC consumer, similar to Diamond's, which it will price beginning at $499.

Wired, Inc. introduced free DVD playback software for its MasonX PowerMac MPEG board. The software allows the card to play DVD movies from either computer DVD-ROM drives, from DVD files stored on a hard disk, from a standard CD-ROM drive, or from a computer network. While the software is free, the card costs $899.

While title producers will focus on movies and multimedia, some companies, like **SilverPlatter**, are poised to convert some of their titles to the new format. SilverPlatter demonstrated its first database on DVD-ROM, **The Union Catalogue of Belgian Research Libraries**, at the **Online Information 96 conference** held in London last December. It plans to follow it with **MEDLINE Advanced** by mid-1997 and other databases later in the year.

At present, DVD looks like an expensive, nonrecordable alternative to the VCR. While recordable and erasable DVD drives could arrive before the end of the year, they still won't have the re-recordable flexibility of a VCR.

Just as the audio CD was adapted for text and graphics, DVD-ROM could find a welcome audience in the library. It will be particularly appropriate for multimedia-rich titles. Titles could also be written with HTML code to permit downloading updates straight from the Web. The large storage capacity could solve the problems of handling multi-disc titles or large collections of images. Whether it will support a common user interface or be more user friendly than current systems remains to be seen.

It appears that DVD-ROM will eventually supplant CD-ROM; but the full transition will probably take a couple of years. With drives priced around $800 for the forseeable future, adoption will be slow. The issues related to networking the hardware and the development of DVD-ROM jukeboxes will also need to be addressed. In any case, adoption will require libraries to replace their hardware infrastructures. Whether or not DVD-ROM will provide enough incentive to do so or to break the popular fascination with the Internet remains to be seen.